

THE EVOLUTION OF METALLICITY AND METALLICITY GRADIENTS FROM $Z = 2.7 - 0.6$ WITH KMOS^{3D}

EVA WUYTS¹, EMILY WISNIOSKI¹, MATTEO FOSSATI^{2,1}, NATASCHA M. FÖRSTER SCHREIBER¹, REINHARD GENZEL^{1,3,4}, RIC DAVIES¹, J. TREVOR MENDEL¹, THORSTEN NAAB⁵, BERNHARD RÖTTGERS⁵, DAVID J. WILMAN^{2,1}, STIJN WUYTS⁶, KAUSHALA BANDARA¹, ALESSANDRA BEIFIORI^{2,1}, SIRIO BELLI¹, RALF BENDER^{1,2}, GABRIEL B. BRAMMER⁶, ANDREAS BURKERT⁵, JEFFREY CHAN^{2,1}, AUDREY GALAMETZ¹, SANDESH K. KULKARNI¹, PHILIPP LANG¹, DIETER LUTZ¹, IVELINA G. MOMCHEVA⁷, ERICA J. NELSON⁸, DAVID ROSARIO¹, ROBERTO P. SAGLIA^{1,2}, STELLA SEITZ², LINDA J. TACCONI¹, KEN-ICHI TADAKI¹, HANNAH ÜBLER¹, PIETER VAN DOKKUM^{8*}

Draft version March 4, 2016

ABSTRACT

We present measurements of the $[\text{N II}]/\text{H}\alpha$ ratio as a probe of gas-phase oxygen abundance for a sample of 419 star-forming galaxies at $z = 0.6 - 2.7$ from the KMOS^{3D} near-IR multi-IFU survey. The mass-metallicity relation (MZR) is determined consistently with the same sample selection, metallicity tracer, and methodology over the wide redshift range probed by the survey. We find good agreement with long-slit surveys in the literature, except for the low-mass slope of the relation at $z \sim 2.3$, where this sample is less biased than previous samples based on optical spectroscopic redshifts. In this regime we measure a steeper slope than some literature results. Excluding the AGN contribution from the MZR reduces sensitivity at the high mass end, but produces otherwise consistent results. There is no significant dependence of the $[\text{N II}]/\text{H}\alpha$ ratio on SFR or environment at fixed redshift and stellar mass. The IFU data allow spatially resolved measurements of $[\text{N II}]/\text{H}\alpha$, from which we can infer abundance gradients for 180 galaxies, thus tripling the current sample in the literature. The observed gradients are on average flat, with only 15 gradients statistically offset from zero at $> 3\sigma$. We have modelled the effect of beam-smearing, assuming a smooth intrinsic radial gradient and known seeing, inclination and effective radius for each galaxy. Our seeing-limited observations can recover up to 70% of the intrinsic gradient for the largest, face-on disks, but only 30% for the smaller, more inclined galaxies. We do not find significant trends between observed or corrected gradients and any stellar population, dynamical or structural galaxy parameters, mostly in agreement with existing studies with much smaller sample sizes. In cosmological simulations, strong feedback is generally required to produce flat gradients at high redshift.

Subject headings: galaxies: high-redshift, galaxies: abundances, galaxies: evolution

1. INTRODUCTION

The evolution of the cosmic star formation density and the mass growth of galaxies is largely driven by their available gas reservoirs, as determined from the interplay between gas accretion through cosmic inflow and mergers, star formation, and gas outflows driven by AGN and stellar feedback (e.g. Genel et al. 2008; Dekel et al. 2009; Bouché et al. 2010; Davé et al. 2012; Lilly et

al. 2013; Tacconi et al. 2013). The same processes also determine the metal content of galaxies, its spatial distribution and its evolution over cosmic time. Observationally, we find a tight correlation between stellar mass and gas-phase oxygen abundance in the local Universe, as robustly quantified in the Sloan Digital Sky Survey (SDSS Tremonti et al. 2004). Theoretical models explain this relation through a combination of momentum- or energy-driven winds to remove metal-rich gas from the galaxy, inflows of metal-poor gas from the surrounding intergalactic medium, and variations in the star formation efficiency of galaxies (e.g. Dalcanton et al. 2004; Brooks et al. 2007; Finlator & Davé 2008; Spitoni et al. 2010; Peeples et al. 2011; Lu et al. 2015). Observational constraints on the mass-metallicity relation (MZR) and its evolution over cosmic time thus provide useful benchmarks for semi-analytic and numerical galaxy evolution models addressing the relative importance of gas inflow, enrichment, recycling and outflow.

Towards higher redshift, an overall decrease in metallicity at fixed stellar mass has been well established with large samples since the recent developments of multi-object spectrographs and the availability of grism data from the Hubble Space Telescope (e.g. Erb et al. 2006; Zahid et al. 2011; Henry et al. 2013; Stott et al. 2013; Cullen et al. 2014; Steidel et al. 2014; Wuyts, E. et al. 2014; Zahid et al. 2014; Sanders et al. 2015). In absolute terms, the normalisation of the MZR evolution

¹ Max-Planck-Institut für extraterrestrische Physik, Giessenbachstr. 1, D-85741 Garching, Germany (evawuyts@mpe.mpg.de)

² Universitäts-Sternwarte München, Scheinerstr. 1, München, D-81679, Germany

³ Department of Physics, Le Conte Hall, University of California, 94720 Berkeley, USA

⁴ Department of Astronomy, Hearst Field Annex, University of California, Berkeley, 94720, USA

⁵ Max-Planck Institute for Astrophysics, Karl Schwarzschildstrasse 1, D-85748 Garching, Germany

⁶ Department of Physics, University of Bath, Claverton Down, Bath, BA2 7AY, UK

⁷ Space Telescope Science Institute, Baltimore, MD 21218, USA

⁸ Department of Astronomy, Yale University, P.O. Box 208101, New Haven, CT 06520-810, USA

* Based on observations obtained at the Very Large Telescope (VLT) of the European Southern Observatory (ESO), Paranal, Chile (ESO program IDs 092.A-0091, 093.A-0079, 094.A-0217, and 095.A-0047). This work is further based on observations taken by the 3D-HST Treasury Program (GO 12177 and 12328) with the NASA/ESA Hubble Space Telescope, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555.

remains uncertain due to its dependence on the metallicity indicator used (Kewley & Ellison 2008). Since the set of observable rest-frame optical emission lines shifts with redshift, studies targeting different redshift ranges typically use different strong-line indicators, which complicates a direct comparison. Additionally, it remains unclear whether for the same indicator, the locally derived calibration remains valid at higher redshift given the different photo-ionisation conditions of high- z galaxies (e.g. Kewley et al. 2013; Steidel et al. 2014; Shapley et al. 2015). New approaches using Bayesian analysis or χ^2 minimisation for a larger set of emission lines to jointly solve for the galaxy metallicity as well as interstellar medium (ISM) properties such as ionisation parameter and electron density, are attempting to address this (Pérez-Montero 2014; Blanc et al. 2015).

In addition to the total metal content of galaxies, the spatial distribution of heavy elements within galaxies further constrains their baryonic and chemical assembly history. In the local Universe, abundance measurements of individual H II regions generally find that galaxies have negative gradients, where the inner regions have higher metallicities than the outer disk regions, indicative of an inside-out growth scenario (Zaritsky et al. 1994; van Zee et al. 1998; Sánchez et al. 2014). There are indications of a flattening of the gradient outside the isophotal radius, possibly due to lower star formation efficiency in the outer disk (e.g. Goddard et al. 2011; Bresolin et al. 2012). Mergers have been found to exhibit flat gradients, likely due to interaction-induced inflow of metal-poor gas in the centre and metal mixing (Kewley et al. 2010; Rupke et al. 2010a,b; Perez et al. 2011; Sánchez et al. 2014). At high redshift, dilution from gas infall into the galaxy centre has been invoked to explain some observations of positive abundance gradients (Cresci et al. 2010).

The time evolution of abundance gradients can be predicted from classical analytical models of the chemical evolution of the Milky Way disk (Chiappini et al. 2001; Mollá & Díaz 2005; Fu et al. 2009), or from cosmological hydrodynamical simulations (Kobayashi & Nakasato 2011; Rahimi et al. 2011; Few et al. 2012; Pilkington et al. 2012; Gibson et al. 2013; Anglés-Alcázar et al. 2014). Pilkington et al. (2012) find little agreement in an extensive comparison of different simulations and chemical evolution models, which they trace back to the varying treatment of star formation and feedback in the simulations. The impact of the assumed feedback scenario is examined further by Gibson et al. (2013). In this study, “normal” feedback (using 10-40% of SN energy to heat the surrounding ISM) predicts steep gradients of -0.3 dex/kpc at $z \sim 2$ which flatten over time, while “enhanced” feedback (including radiation pressure from massive stars) can better redistribute metal-enriched gas over large spatial scales and produces relatively flat gradients at all times. These studies highlight how robust gradient measurements at high redshift can provide powerful constraints on the uncertain nature of stellar feedback processes.

Unfortunately also the observational measurements present challenges, requiring significant time investments to spatially resolve the fainter emission lines (typically [N II] and/or H β) used in the strong-line metallicity calibrations. The current sample available in the literature

consists of 90 gradients at $z = 0.8 - 3.8$. Adaptive-optics (AO) assisted SINFONI observations have resulted in measured gradients for nine H α -selected HiZELS galaxies at $z = 0.8 - 2.24$ (Swinbank et al. 2012) and 19 SINS/zCSINF galaxies at $z = 1.4 - 2.4$ (Förster Schreiber et al. 2016, in prep.). Without AO, there are 26 gradients at $z \sim 1.2$ from MASSIV (Queyrel et al. 2012) and 10 gradients at $z \sim 3 - 4$ from AMAZE (Troncoso et al. 2014). Taking advantage of the multiplexing capabilities of KMOS at the VLT, Stott et al. (2014) present a first sample of 20 gradients at $z \sim 0.8$ from HiZELS. There are concerns whether the coarse spatial resolution of seeing-limited observations, typically 5 kpc at $z \sim 2$, can recover the intrinsic gradient (Yuan et al. 2013). AO-assisted or space-based observations of lensed galaxies can reach spatial resolutions down to 200-300 pc in the source plane. For 16 lensed galaxies with OSIRIS+AO data (Jones et al. 2010; Yuan et al. 2011; Jones et al. 2013; Leethochawalit et al. 2015) and one arc with HST grism data from the Grism Lens-Amplified Survey from Space (GLASS, Jones et al. 2015), some steeper gradients have been recovered, but most measurements are comparable to non-lensed samples at similar redshift.

In this work, we exploit the large redshift coverage of the KMOS^{3D} survey from $z = 2.7$ to $z = 0.6$ for a consistent study of the evolution of the MZR over five billion years around the period of peak star formation activity in the Universe. We use the spatially resolved information unique to our integral field spectroscopic survey to measure abundance gradients for 180 galaxies, thus tripling the total sample. The Paper is organised as follows. §2 summarises the KMOS^{3D} sample selection, observations, and data reduction. The mass-metallicity relation and its dependence on the presence of an AGN, as well as on SFR and galaxy environment is analysed in §3. In §4 we present the radial abundance gradients, and address the effect of beam-smearing. §5 provides a summary. Throughout this work, we adopt the Chabrier (2003) initial mass function and a flat cosmology with $\Omega_M = 0.3$ and $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$.

2. THE KMOS^{3D} SAMPLE

The galaxies analysed here are taken from the first two years of the KMOS^{3D} survey, which covers observations up to April 2015. The survey is described in detail by Wisnioski et al. (2015). KMOS^{3D} is a five-year GTO survey with the multi-object near-IR integral field spectrograph KMOS at the VLT (Sharples et al. 2013), which aims to observe a mass-selected sample of > 600 galaxies at $z = 0.7 - 2.7$ in H α emission to study their spatially resolved kinematics and star formation. Targets are pre-selected to avoid OH skyline contamination based on a combination of existing spectroscopic campaigns and 3D-HST grism redshifts from the v4 catalogs (Brammer et al. 2012; Momcheva et al. 2015) in the southern CANDELS fields UDS, COSMOS and GOODS-S (Grogin et al. 2011; Koekemoer et al. 2011). The availability of near-IR grism data avoids the bias towards bluer colours intrinsic in any selection based only on optical spectroscopic redshifts. Stellar masses, UV+IR star formation rates and global extinction values for the sample are derived from the CANDELS photometry (Skelton et al. 2014) and available far-IR photometry (Lutz et al. 2011;

Magnelli et al. 2013), using standard spectral energy distribution modelling with FAST (Kriek et al. 2009) for a grid of Bruzual & Charlot (2003) models with exponentially declining star formation histories, Calzetti et al. (2000) dust extinction, and solar metallicity (see Wuyts, S. et al. (2011) for details). Morphological parameters such as inclination and effective radius are available from GALFIT fitting of the F160W CANDELS data by van der Wel et al. (2012).

This work includes 419 targets, 176-63-180 each in the YJ , H and K -bands respectively. The survey strategy deliberately focused on the $z \sim 0.9$ and $z \sim 2.3$ redshift slices first, only now starting to fill in the intermediate range at $z \sim 1.5$ in H -band. Figure 1 situates the sample in the star formation vs stellar mass plane. Our deep integrations (baseline of 4-6-8hrs in YJ , H , and K respectively) allow us to probe significantly below the main-sequence of star formation, though in this parameter space the detection fraction drops below our average of 76%, especially in K -band.

The data were reduced with the Software Package for Astronomical Reduction with KMOS (SPARK, Davies et al. 2013), complemented with custom IDL scripts for additional flat fielding, accurate centering and combining of data taken over multiple epochs, and PSF characterisation. Maps of the $H\alpha$ flux and kinematics are derived with the IDL emission line fitting code LINEFIT, which was originally developed for SINFONI data (Förster Schreiber et al. 2009; Davies et al. 2011). We mask out all spatial pixels where the S/N in $H\alpha$ drops below 5, as well as outliers in velocity or velocity width or in their respective uncertainties. More details on the data reduction and kinematic mapping can be found in Wisnioski et al. (2015).

Guided by the KMOS^{3D} survey design which targets primarily $H\alpha$ emission, we derive gas-phase metallicities from the N2 index, the ratio of $[\text{N II}] \lambda 6583$ to $H\alpha$ emission line fluxes. We note a number of known concerns related to this metallicity indicator: contribution of AGN and shock ionisation to the $[\text{N II}]$ emission (further addressed in § 3.1), saturation at high metallicities, variations in the N/O ratio and secondary nitrogen production, and contribution of the warm diffuse ISM. Given these possible biases, we show our results in terms of the observed $[\text{N II}]/H\alpha$ ratio. When a conversion to metallicity is required for a comparison to the literature, the linear conversion by Pettini & Pagel (2004) is used, $12 + \log(\text{O}/\text{H}) = 8.9 + 0.57 \times \log([\text{N II}]/H\alpha)$. This carries a systematic uncertainty of 0.18 dex.

3. THE MASS-METALLICITY RELATION

We create an integrated 1D spectrum for each galaxy by co-adding all the spatial pixels within the mask after velocity-shifting each spatial pixel to remove the imprint of the overall galaxy kinematics. The $H\alpha$ and $[\text{N II}]$ emission lines are jointly fit with a three-component Gaussian model, forcing a common redshift and line width, and constraining the $[\text{N II}] \lambda 6583/[\text{N II}] \lambda 6548$ doublet ratio to its theoretical value of 3.071 (Storey & Zeippen 2000). Uncertainties are derived via a Monte Carlo approach where the spectrum is perturbed following its noise spectrum. From these fits we derive $H\alpha$ -based SFRs, cor-

rected for dust extinction using the SED-derived reddening and accounting for additional extinction of the nebular lines following Wuyts, S. et al. (2013). The integrated $[\text{N II}]$ emission is detected at $> 3\sigma$ for 90% of the sample, $[\text{N II}]/H\alpha$ ratios are reported in Table 3 in the Appendix.

3.1. AGN contamination

It is well known that the $[\text{N II}]$ emission can be contaminated by the ionising spectrum of an active galactic nucleus. There are a number of different AGN indicators in use in the literature, all of which suffer from biases and incompleteness. The “classical” indicators include X-ray emission, radio emission and mid-IR colours. Keeping in mind that the coverage and depth of the various surveys varies across the CANDELS fields (see Genzel et al. 2014, for more details), we find a typically low percentage of 7% AGN in our KMOS^{3D} sample. Surveys which also target $[\text{O III}]$ and $H\beta$ emission can use the optical BPT diagnostic diagram (Baldwin et al. 1981) to identify the presence of AGN in their sample, and typically find another $\sim 10\%$ contamination (Zahid et al. 2014; Sanders et al. 2015). The KMOS^{3D} survey currently does not cover the $[\text{O III}]$ or $H\beta$ line for the majority of its targets, but $[\text{N II}]/H\alpha$ alone also serves as a useful diagnostic to identify AGN from rest-frame optical emission lines at high redshift (Stasińska et al. 2006; Cid Fernandes et al. 2011; Coil et al. 2015). We find that 8% of targets in our sample lie above a threshold $[\text{N II}]/H\alpha > 0.6$, using the threshold from Kewley et al. (2001).

The spatially resolved IFU data give access to two additional AGN indicators. The AGN contribution to the line emission is expected to be largest in the nuclear region of the galaxy, and possibly diluted by the emission from the outer disk. We therefore create a nuclear spectrum for each of our targets by co-adding only the central $0.4''$. The nuclear $[\text{N II}]/H\alpha$ ratio lies above 0.6 for an additional 5% of the sample. Secondly, Förster Schreiber et al. (2014) and Genzel et al. (2014) reported evidence for nuclear AGN-driven outflows of ionised gas based on the presence of a broad underlying component with FWHM=500-2000 km/s in the line profile of $H\alpha$ and the forbidden lines. We visually examine all nuclear spectra as described by Genzel et al. (2014) and find evidence of underlying broad emission for 27% of the sample analysed in this paper. The incidence of each of the AGN indicators described above is a strong function of stellar mass, such that at $\log(M_*/M_\odot) > 10.9$, 70% of the sample satisfies one or more of the AGN indicators.

3.2. Stacking

The KMOS^{3D} sample has a high detection rate of 90% for the galaxy-integrated $[\text{N II}]$ emission, not unsurprising given the required depth to spatially resolve the $H\alpha$ emission. Nevertheless, we stack our galaxies in bins of stellar mass to robustly take into account the upper limits. Similarly to Wuyts, E. et al. (2014), we use inverse variance weighting of continuum-subtracted, velocity-corrected and $H\alpha$ -normalized integrated spectra. The line fluxes and uncertainties are derived as the jackknife mean and standard error and reported in Table 1. We aim for 15-20 targets per bin. Figure 2 shows the mass-metallicity relation for the stacked spectra with and without the AGN identified in the previous section. The strong dependence of AGN incidence

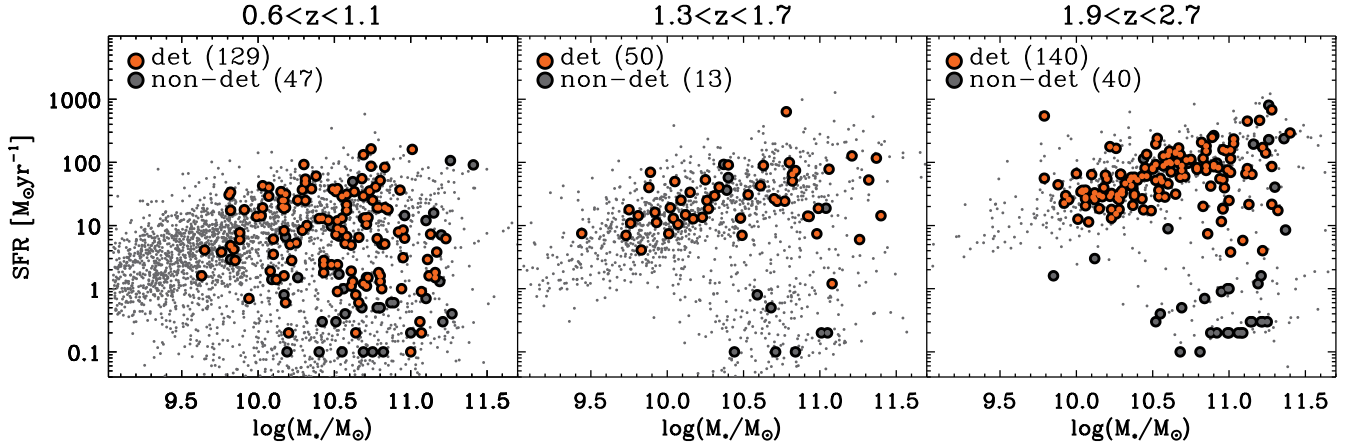


FIG. 1.— Location of the observed KMOS^{3D} sample in the SED-derived SFR vs stellar mass plane in three redshift bins. The underlying grey dots represent the full mass-selected galaxy population at $z = 0.6 - 2.7$ in the CANDELS fields. Our deep observations probe significantly below the main-sequence of star formation.

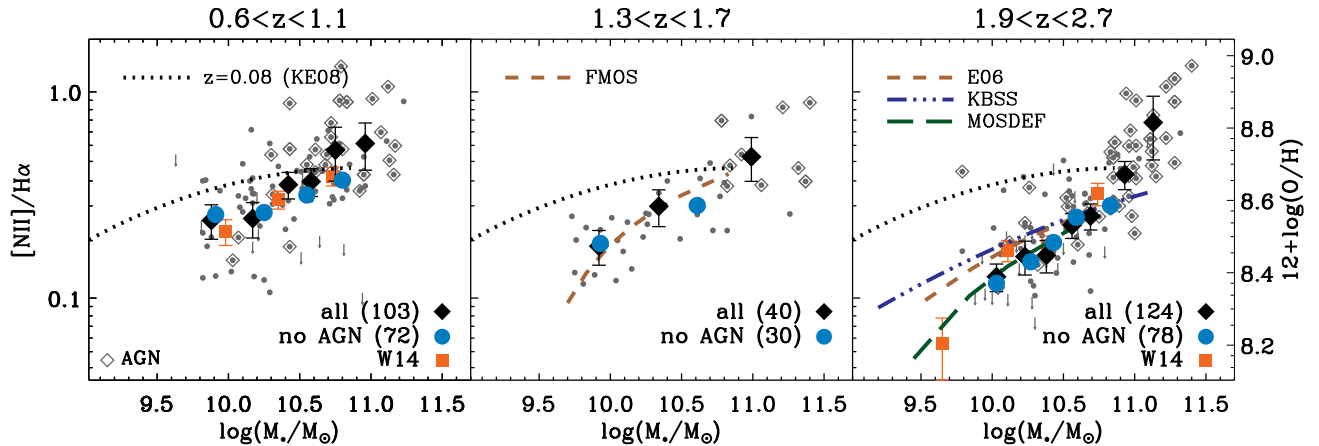


FIG. 2.— Galaxy-integrated $[\text{N II}]/\text{H}\alpha$ ratios as a function of stellar mass for the KMOS^{3D} sample. The right axis expresses galaxy abundance as $12 + \log(\text{O}/\text{H})$ as derived from the linear relation by Pettini & Pagel (2004). The grey datapoints in the background show our individual detections and upper limits, AGN are surrounded by an open diamond. The large black diamonds and blue circles correspond to the stacked spectra with and without AGN-identified targets respectively. The local relation from Kewley & Ellison (2008) is shown with a dotted line in each panel. The orange squares correspond to the earlier KMOS+SINFONI+LUCI results from Wuyts, E. et al. (2014). Other studies in the literature include the FMOS survey at $z \sim 1.6$ (Zahid et al. 2014, brown dashed), and at $z \sim 2.3$ the sample from Erb et al. (2006) (brown dashed), the KBSS survey (Steidel et al. 2014, dark blue dot dashed) and the MOSDEF survey (Sanders et al. 2015, dark green dashed). Their MZR relations are represented consistently based on metallicities derived from the N2 indicator and the linear Pettini & Pagel (2004) conversion.

on stellar mass results in a loss of many of the highest mass galaxies. However, in the overlapping mass regime at $\log(M_*/M_\odot) \lesssim 10.8$, both stacked relations are consistent within the uncertainties, suggesting that in this mass range the effect of the AGN contribution to the galaxy-integrated $[\text{N II}]$ emission is limited. At higher masses, the integrated $[\text{N II}]/\text{H}\alpha$ ratios are often biased high.

The results are in excellent agreement with the MZR presented in Wuyts, E. et al. (2014) based on the first year KMOS^{3D} data, augmented with observations with SINFONI at the VLT and the long-slit near-IR spectrograph LUCI at the Large Binocular Telescope on Mount Graham, Arizona. The latter two datasets are responsible for the extension to lower stellar masses, especially in K -band. Further comparison to the literature at $z \sim 1.6$ is available from the FMOS survey, for 168 galaxies in

the COSMOS field observed with the multi-object near-IR spectrograph FMOS on Subaru (Zahid et al. 2014; Silverman et al. 2015). The wide field of view surveyed includes many more very massive targets compared to KMOS^{3D}, but the lack of spatially-resolved data limits the detection of AGN contributions, which could explain the somewhat higher abundances at the high mass end, where we have shown AGN to bias the $[\text{N II}]/\text{H}\alpha$ ratios high. Most high- z emission-line science has so far been carried out at $z \sim 2.3$, where the full suite of rest-frame optical diagnostic lines can be observed in the near-IR wavelength range. The MZR presented by Erb et al. (2006) was long the only study with a large sample (~ 100) at this redshift, albeit with few individual $[\text{N II}]$ detections. Recently, the new multi-object near-IR spectrograph MOSFIRE on Keck has changed the landscape with two large emission-line surveys of many hundreds of galaxies each. The MOSFIRE Deep Evolution

TABLE 1
STACKED MASS-METALLICITY RELATION IN THREE REDSHIFT
INTERVALS, WITH AND WITHOUT AGN.

	all		no AGN	
Redshift	$\log(M_*/M_\odot)$	$[\text{N II}]/\text{H}\alpha$	$\log(M_*/M_\odot)$	$[\text{N II}]/\text{H}\alpha$
$z \sim 0.9$	9.88	0.186 ± 0.026	9.91	0.218 ± 0.043
	10.17	0.250 ± 0.065	10.25	0.277 ± 0.044
	10.42	0.338 ± 0.068	10.55	0.364 ± 0.072
	10.58	0.436 ± 0.049	10.80	0.384 ± 0.038
	10.75	0.404 ± 0.067		
	10.96	0.432 ± 0.122		
$z \sim 1.5$	9.92	0.165 ± 0.026	9.93	0.164 ± 0.034
	10.34	0.233 ± 0.040	10.61	0.262 ± 0.057
	10.99	0.368 ± 0.204		
$z \sim 2.3$	10.03	0.119 ± 0.016	10.03	0.114 ± 0.012
	10.23	0.122 ± 0.026	10.27	0.134 ± 0.024
	10.38	0.142 ± 0.017	10.43	0.165 ± 0.020
	10.56	0.211 ± 0.032	10.59	0.231 ± 0.034
	10.69	0.293 ± 0.062	10.83	0.328 ± 0.115
	10.93	0.351 ± 0.049		
	11.13	0.542 ± 0.400		

Field (MOSDEF) survey follows a similar target selection to KMOS^{3D} in the 3D-HST CANDELS fields observable from Mauna Kea (Kriek et al. 2015). The Keck Baryonic Structure Survey (KBSS) targets 15 quasar fields with a large sample of spectroscopically confirmed redshifts at $1.5 \lesssim z \lesssim 3.5$ and a wealth of multi-wavelength ancillary data (e.g. Steidel et al. 2004; Rudie et al. 2012; Steidel et al. 2014). We have taken care to consistently compare relations which have been derived with the same N2 indicator and linear Pettini & Pagel (2004) conversion employed here. Our results are in perfect agreement with the MOSDEF relation (Sanders et al. 2015), which is reassuring given our almost identical parent sample. The MZR relation reported by KBSS has a shallower slope, which is likely due to differences in the selected target population (Strom et al. 2016, in prep.).

3.3. The Effect of Star Formation Rate

In the local Universe, Ellison et al. (2008) first reported an anti-correlation between metallicity and specific star formation rate (sSFR). Further studies showed how including the SFR as a second parameter in the mass-metallicity relation reduces the scatter (Mannucci et al. 2010; Andrews & Martini 2013). There have been claims that this “fundamental” relation (FMR) between stellar mass, SFR and galaxy abundance is independent of redshift, such that the lower abundances observed at high redshift at fixed stellar mass are fully explained by the overall higher SFR at these epochs (e.g. Mannucci et al. 2010; Belli et al. 2013; Henry et al. 2013; Stott et al. 2013). However, the majority of high- z studies report no dependence of the MZR on star formation rate, or an offset from the local FMR (Zahid et al. 2014; Wuyts, E. et al. 2014; Steidel et al. 2014; Troncoso et al. 2014; Sanders et al. 2015; Grasshorn Gebhardt et al. 2015).

Our data does not show a significant trend between inferred metallicity and sSFR, in contrast to recent results from the KMOS Redshift One Spectroscopic Survey (KROSS Magdis et al. 2016). This could be due to a less complete removal of AGN contamination from their

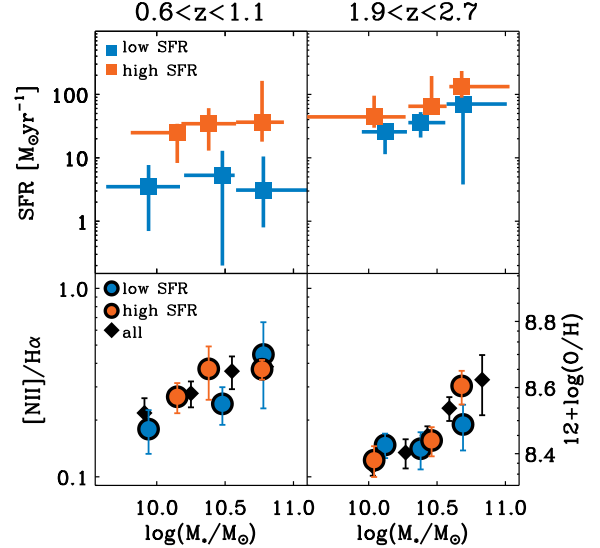


FIG. 3.— The effect of star formation on metallicity, using the dust-corrected SFRs derived from the integrated H α flux. Top panels show the median value and range of a low and high SFR bin for three stellar mass bins at $z \sim 0.9$ and $z \sim 2.3$. The MZR for the low and high SFR bins separately is shown in the bottom panels, and compared to the MZR for all galaxies from Figure 2. There is no significant offset between the SFR bins.

sample as well as a difference in survey depth at $z \sim 1$ (KMOS^{3D}: ~ 5 hrs; KROSS: ~ 2.5 hrs). To investigate the dependence on SFR at fixed mass and redshift, we split our AGN-excluded sample at $z \sim 0.9$ and $z \sim 2.3$ in two bins of SFR, using the dust-corrected SFRs derived from the integrated H α flux (Figure 3). The intermediate redshift bin at $z \sim 1.5$ does not yet contain sufficient targets. We do not see a significant offset between the MZR separately derived for the low and high SFR bin. This does not change when we use the photometrically based SFR estimates from UV+IR broad-band data. The dynamic range of the $z \sim 2.3$ sample is very limited, but over the SFR range probed at $z \sim 0.9$ in KMOS^{3D}, a significant effect on metallicity is seen in the local Universe (Andrews & Martini 2013).

The equilibrium or “gas-regulator” picture of galaxy evolution driven by the balance between gas accretion, star formation and gas outflow, naturally finds a fundamental relation between stellar mass, SFR and metallicity (e.g. Lilly et al. 2013). The lack of observed correlation remains puzzling, and could signify departures from the first-order assumptions of these simple analytical equilibrium models. First, however, the possible caveats related to metallicity measurements based on strong-line indicators should be more completely understood.

3.4. The Effect of Environment

Local density can have a strong effect on the overall evolution of galaxies. Specifically for the galaxy abundance, a denser environment could remove metal-poor gas from the outskirts of galaxies or result in shorter gas recycling times (Oppenheimer & Davé 2008), enhancing the observed metallicity. However, offsets in the local

MZR of field and cluster star-forming galaxies are observationally and theoretically constrained to < 0.05 dex (Ellison et al. 2009; Davé et al. 2011; Scudder et al. 2012; Hughes et al. 2013). The few studies carried out to date in proto-cluster environments at $z \sim 2$ mostly limit any enhancement of the abundance to ≤ 0.15 dex at $< 2\sigma$ significance (Kulas et al. 2013; Shimakawa et al. 2015; Kacprzak et al. 2015); Valentino et al. (2015) in contrast find a metal deficiency of 0.25 dex at 4σ significance for a cluster at $z = 1.99$.

With the KMOS^{3D} sample we can for the first time probe the effect of variations in local density on galaxy abundances for a large sample at high redshift. We characterise each KMOS^{3D} target as a central or satellite galaxy based on a determination of its mass-rank in an adaptive aperture and a comparison to simulations (Fossati et al. 2015, Fossati et al. 2016 in prep.). The expected $\sim 30\%$ satellite galaxies is roughly recovered, except at $z \sim 2.3$ due to the stronger focus on massive galaxies in this redshift slice in the first two years of KMOS^{3D} target selection. Central and satellite galaxies occupy a similar phase space in the SFR - stellar mass plane, as shown in the top panel of Figure 4. At $z \sim 0.9$, the sample contains enough non-AGN, [N II] detected satellite galaxies to stack both categories separately in bins of stellar mass and compare the MZR. We find a metallicity enhancement of 0.06 dex for central galaxies, though at low significance ($< 2\sigma$) given the uncertainties.

4. ABUNDANCE GRADIENTS

The spatial distribution of the [N II]/H α ratio across individual galaxies contains additional information beyond a simple measurement of the galaxy-integrated abundance. For each target, a set of elliptical annuli is created, centred on the continuum light, angled along the kinematic position angle (see Wisnioski et al. 2015) and with an ellipticity matching the outer H α contours. Given the FWHM of our seeing-limited data, we ideally choose apertures with a width of $0.4''$, or 2 spatial pixels. However, only 40% of the sample has enough signal-to-noise over a large enough area to detect [N II] at $> 3\sigma$ in three subsequent 2 pixel-wide apertures, which is our requirement to robustly measure a radial gradient. Especially in the highest redshift bin at $z \sim 2.3$, most galaxies are too small and for only 17% can a gradient be measured. Therefore, our default apertures have a width of $0.2''$. Where possible, we have checked that gradients measured with $0.2''$ and $0.4''$ wide apertures are consistent within the uncertainties. In each aperture, we create a velocity-corrected integrated spectrum and jointly fit the [N II] and H α emission as described in Section 3. A linear fit of the [N II]/H α ratio as a function of the semi-major axis radius (defined in the middle of each annulus) results in a measurement of the gradient $\Delta N2/\Delta r$ in dex/kpc. Figure 11 in the Appendix shows the H α , velocity and [N II]/H α map as well as the [N II]/H α ratio as a function of radius for five example galaxies to illustrate the process. We have visually examined each gradient measurement to avoid catastrophic failures in the fit.

For the galaxies identified as having some AGN contribution from one of the indicators described in Section 3.1, the [N II] emission in the central region could be enhanced by the AGN and thus not accurately represent the

metallicity of the ionised gas. In these cases, we exclude the inner two apertures (i.e. the inner $0.4''$) from the radial fit. When only four apertures with robust [N II] detections exist, only the innermost aperture is excluded. For these targets, removing two apertures and fitting a gradient to only the remaining outer two radial bins gives consistent results. AGN with only three radial [N II] detections are excluded. For the full sample of AGN, the restricted gradients are on average 0.001 ± 0.02 dex/kpc flatter, i.e. fully consistent with the full radial gradients, as can be seen in Figure 5.

The final gradient measurements are reported in Table 3 in the Appendix. Most of the observed gradients are flat, only 15/180 targets exhibit a gradient significantly offset from zero (at $> 3\sigma$), of which 13 gradients are negative and 2 positive. Figure 6 shows the full sample of measured gradients as a function of stellar mass in [dex/kpc] and [dex/ r_e] in the top and bottom panels respectively, color-coded by redshift. Based on a sequence of 1000 Spearman correlation tests, where we randomly vary the sample of gradients according to their uncertainties, the negative correlation with stellar mass is significant at 2.8σ . We find a positive correlation with specific star formation rate (sSFR) at 2.5σ , the significance reduces to 1.5σ when the sSFR is normalised to the main-sequence at each redshift. Stott et al. (2014) have similarly reported a weak positive trend with specific star formation rate at 2.9σ . No significant correlations are present with integrated [N II]/H α or any of the dynamical (dispersion, velocity gradient, v_{obs}/σ_0) or structural (effective radius, inclination) properties of the galaxies.

The somewhat steeper gradients in more massive galaxies could be due to residual influence of AGN- or shock-ionisation on the [N II] emission, since these processes are more prevalent at the high mass end. It could additionally signify stronger inside-out growth in massive galaxies (e.g. Nelson et al. 2015). Finally, feedback can more easily redistribute metals in the shallower potential wells of lower mass galaxies, flattening the gradients. Stott et al. (2014) link the correlation with sSFR to the inflow of metal-poor gas flattening the gradients of strongly star-forming systems.

4.1. Beam-smearing

Before physically interpreting the observed gradients, it is crucially important to understand the effects of the coarse spatial resolution of seeing-limited data on the intrinsic abundance gradients of high-redshift galaxies. For this purpose we have analysed a set of exponential disk models created with the IDL code DYSMAL (Cresci et al. 2009; Davies et al. 2011) for a range of inclination and effective radius. We add [N II] emission assuming a smooth linear radial gradient with a chosen central [N II]/H α ratio and slope. The model cubes are convolved with a Gaussian PSF for a range of assumed FWHM from $0.1''$ for some AO-assisted data to $1.0''$ for seeing-limited data taken in bad observing conditions, and random noise is added to achieve a pre-defined signal-to-noise ratio for the central pixel. Table 2 summarises the default value (in bold) and explored range for each parameter, chosen to reflect the range of properties found in the KMOS^{3D} sample. All model cubes are analysed in the exact same

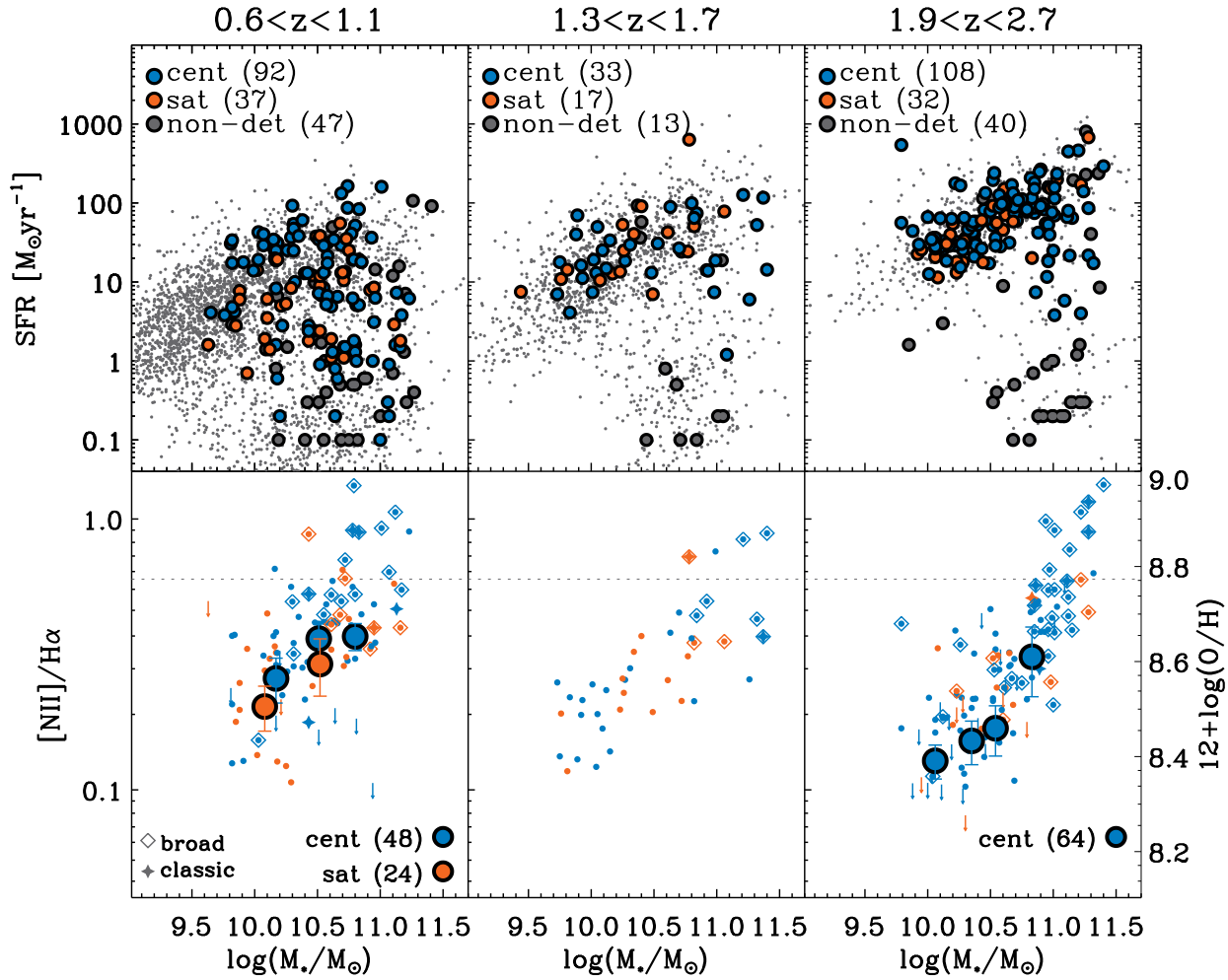


FIG. 4.— (*Top*) SFR vs stellar mass for the KMOS^{3D} sample (as in Figure 1), color-coded as central and satellite galaxies in blue and orange respectively. We find the expected $\sim 30\%$ of satellites (Fossati et al. 2015) at $z \sim 0.9$ and $z \sim 1.5$, at $z \sim 2.3$ the focus on massive galaxies results in fewer satellite galaxies. Both subsets occupy a similar phase space in this diagram. (*Bottom*) Galaxy-integrated $[\text{N II}]/\text{H}\alpha$ ratios vs stellar mass (as in Figure 2), color-coded as central and satellite galaxies in blue and orange respectively. Only at $z \sim 0.9$ do we have enough satellites for a stacked mass-metallicity relation excluding AGN. Central galaxies exhibit marginally higher $[\text{N II}]/\text{H}\alpha$ ratios at 0.06 dex.

TABLE 2
PARAMETERS USED FOR MODEL DATA CUBES TO EXPLORE
BEAM-SMEARING EFFECTS.

FWHM ["]	0.1, 0.2, 0.3, 0.4, 0.5, 0.6 , 0.7, 0.8, 0.9, 1.0 + AO
r_e ["]	0.3, 0.4, 0.5 , 0.6, 0.7, 0.8, 0.9, 1.0
inclination [°]	0, 30, 45 , 60, 75
S/N _{central}	10, 20 , 30
$[\text{N II}]/\text{H}\alpha_{\text{central}}$	0.3, 0.5 , 1.0
gradient [dex/kpc]	0.2, 0.1, 0.05, -0.05, -0.1 , -0.2

way as described in §4 for the KMOS data.

From this first order analysis, we find that the beam-smearing mainly depends on the ratio of the FWHM of the data to the effective radius of the galaxy, which determines how well it can be resolved radially. The right panel of Figure 7 additionally shows that more highly-inclined galaxies and intrinsically steeper gradients are each affected more, as would be expected. The central $[\text{N II}]/\text{H}\alpha$ ratio does not matter, as long as one has

enough sensitivity to detect $[\text{N II}]$ in the outer disk. And finally, the S/N of the central pixel only impacts the uncertainties, as was also found by Yuan et al. (2013).

The KMOS^{3D} data are taken under mostly uniform seeing conditions ($\text{FWHM} = 0.55'' \pm 0.09''$). From our models, we see that we can recover at most 70% of the intrinsic gradient for the largest face-on disks, down to only $\sim 25\%$ in the smallest, highly-inclined targets. To compare beam-smearing effects with AO-assisted datasets, we also ran simulations assuming the PSF properties of SINFONI+AO observations, for which the PSF is best characterised with a double-component Gaussian model to capture the AO-corrected narrow core component with $\text{FWHM} = 0.16''$ and 86% of the peak flux, and the uncorrected broad halo underneath with $\text{FWHM} = 0.5''$ and 14% of the peak flux (Förster Schreiber et al. 2009, Förster Schreiber et al. 2016, in prep.). The halo causes a larger beam-smearing effect than measured for a single Gaussian FWHM of $0.16''$. On average, AO data recovers 65% of the gradient, as opposed to 50% at seeing-limited resolution. Utilizing the HST grism capabilities

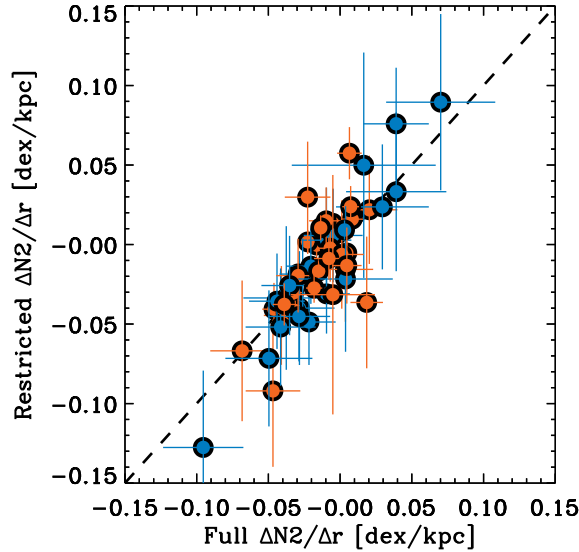


FIG. 5.— A comparison of the full radial gradient with the gradient restricted to the outer disk for the galaxies flagged as AGN. For the orange symbols, the inner $0.4''$ covered by the two innermost apertures is excluded from the fit, for the blue symbols we can exclude only the innermost $0.2''$ to keep at least three radial $[\text{N II}]$ detections. The uncertainties for the restricted gradient on the y-axis are naturally larger due to the fewer radial data points, but on average the AGN does not significantly steepen the derived linear gradient.

as in Jones et al. (2015), delivers a single narrow PSF at $\text{FWHM} \sim 0.2''$ and allows the most accurate gradient measurement.

Two previous efforts to take into account the effects of beam-smearing have been made. Yuan et al. (2013) simulate high- z datacubes with different seeing and S/N values based on the measured central abundance and abundance gradient for one specific local galaxy (IRAS F17222-5953) and recover only 13% of the intrinsic gradient when using three uniform radial bins at $\text{FWHM}=0.6''$. For the KMOS observations of 20 HiZELS galaxies at $z = 0.8$, Stott et al. (2014) perform a simulation of 1000 disk galaxies with a range of inclinations and input gradients and recover 80% of the intrinsic gradient in a face-on disk at $0.7''$ seeing. For an inclination of 80° , this reduces to 30%. This is in contrast to our results, which only recover 35% at $\text{FWHM}=0.7''$ and are much less dependent on inclination. It remains unclear from where this difference arises. From our analysis, a large dependence on inclination is unexpected, since the use of elliptical apertures along the major axis following the axis ratio of the $\text{H}\alpha$ contours is designed to mostly take this into account.

Under the assumption that we know the effective radius and inclination of each KMOS^{3D} target from GALFIT modelling of the F160W CANDELS photometry (van der Wel et al. 2012), and that the line emission mostly follows the stellar light as traced by the F160W continuum, we can derive the necessary beam-smearing correction from our set of models. Figure 12 in the Appendix visualises how this is done for three examples. For a known combination of FWHM, inclination and effective radius, a certain intrinsic gradient will be beam-

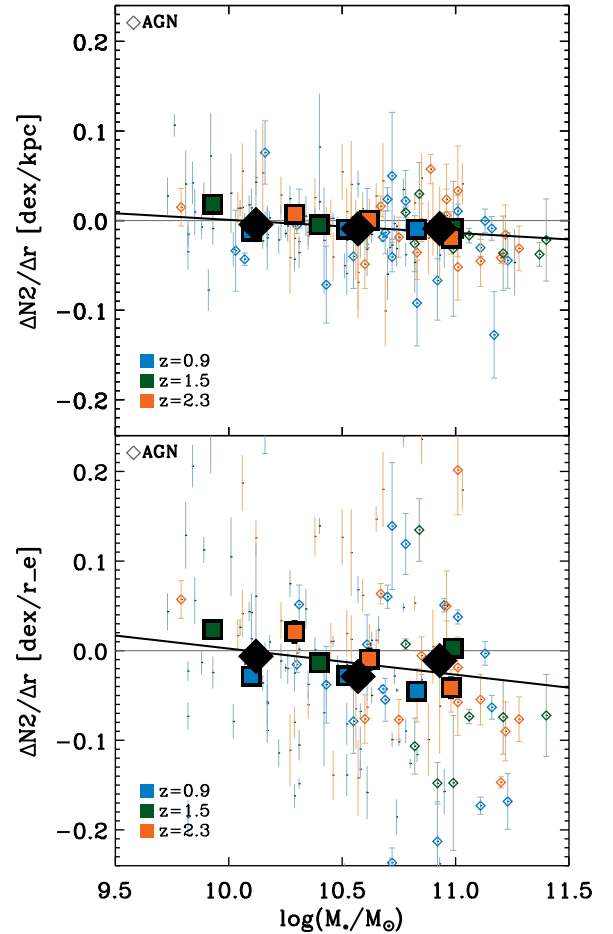


FIG. 6.— Observed metallicity gradients as a function of stellar mass, color-coded by redshift. The *top* panel shows gradients in $[\text{dex/kpc}]$, in the *bottom* panel they have been normalised to the effective radius. Targets surrounded by an open diamond are flagged as AGN by one of the indicators described in § 3.1. The large squares correspond to the weighted average gradient in three bins of stellar mass for each redshift, the black diamonds show the weighted average for the full sample. The black line is a simple 1D fit to those points. From a sequence of 1000 Spearman correlation tests, randomly varying the sample around the uncertainties, this trend is significant at 2.8σ in both panels.

smearred into a unique observed gradient. We derive this match for a range of intrinsic gradients between -0.2 and 0.2 dex/kpc for each target, and then simply invert this correlation to derive the intrinsic gradient that corresponds to the actual observed gradient of the target. A histogram of the resulting beam-smearing correction factors and corrected gradients for the full sample is shown in Figure 8. We caution that each corrected gradient in this histogram has large uncertainties. We re-investigate correlations between the now-intrinsic gradients and the stellar population, structural and kinematic parameters and again do not find any statistically significant trends. Finally, we note that these corrections are derived under the assumption of a smooth, linear radial gradient. Given the clumpy nature and often irregular morphology

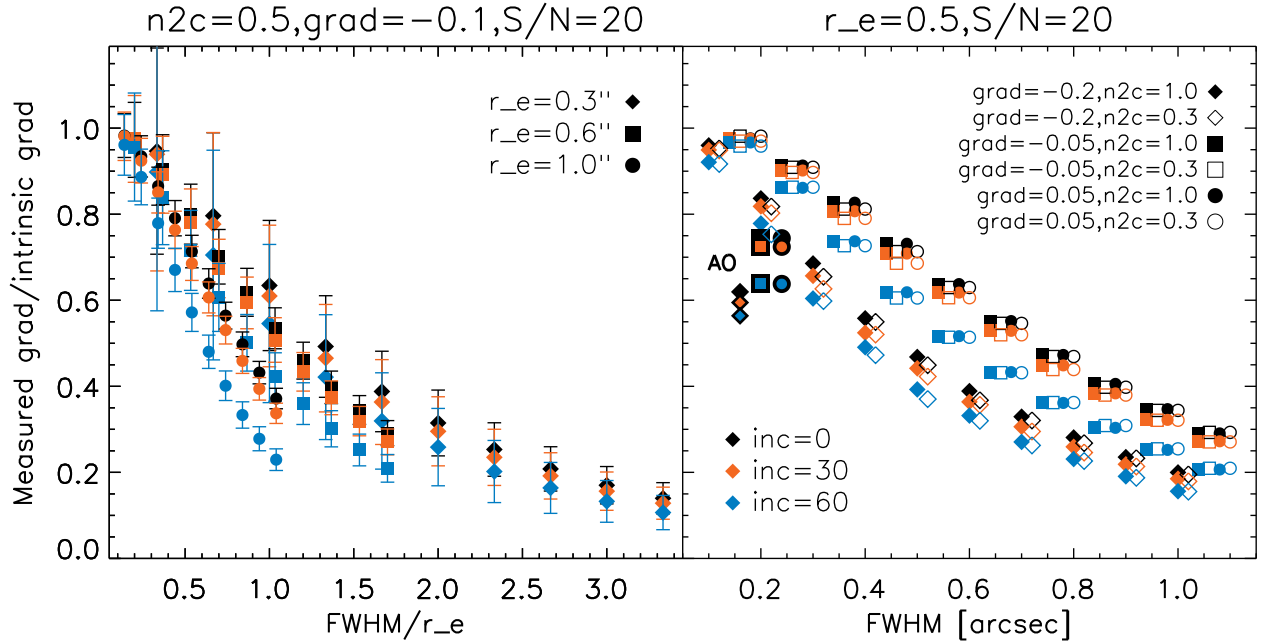


FIG. 7.— Ratio of the recovered vs intrinsic radial $\Delta N_2/\Delta r$ gradient. The color-coding in both panels shows a range of inclinations from face-on (inc=0°) to almost edge-on (inc=60°). In the *left* panel, the x-axis represents the FWHM of the observations divided by the effective radius for $r_e = 0.3, 0.6$ and $1.0''$. The other parameters (central [N II]/H α value (n2c), intrinsic gradient (grad) and S/N) are kept at their default values as specified in Table 2. The ratios have all been derived at the same discretised set of input FWHM, but to improve the clarity of the figure we have shifted the data points for different effective radii slightly along the x-axis. Gradients are more affected by beam-smearing for larger ratios of seeing versus galaxy size. The *right* panel compares the ratio of the recovered vs intrinsic gradient as a function of the FWHM of the observations for different intrinsic gradients and central [N II]/H α ratios. In this panel the effective radius is kept fixed at its default value of $0.5''$. The positions along the x-axis have similarly been shifted to improve the clarity of the figure. A shallower intrinsic gradient is less affected by beam-smearing, while the central N2 ratio has no impact. The black-edged symbols shown at FWHM $\sim 0.16''$ correspond to the double component PSF of SINFONI+AO data.

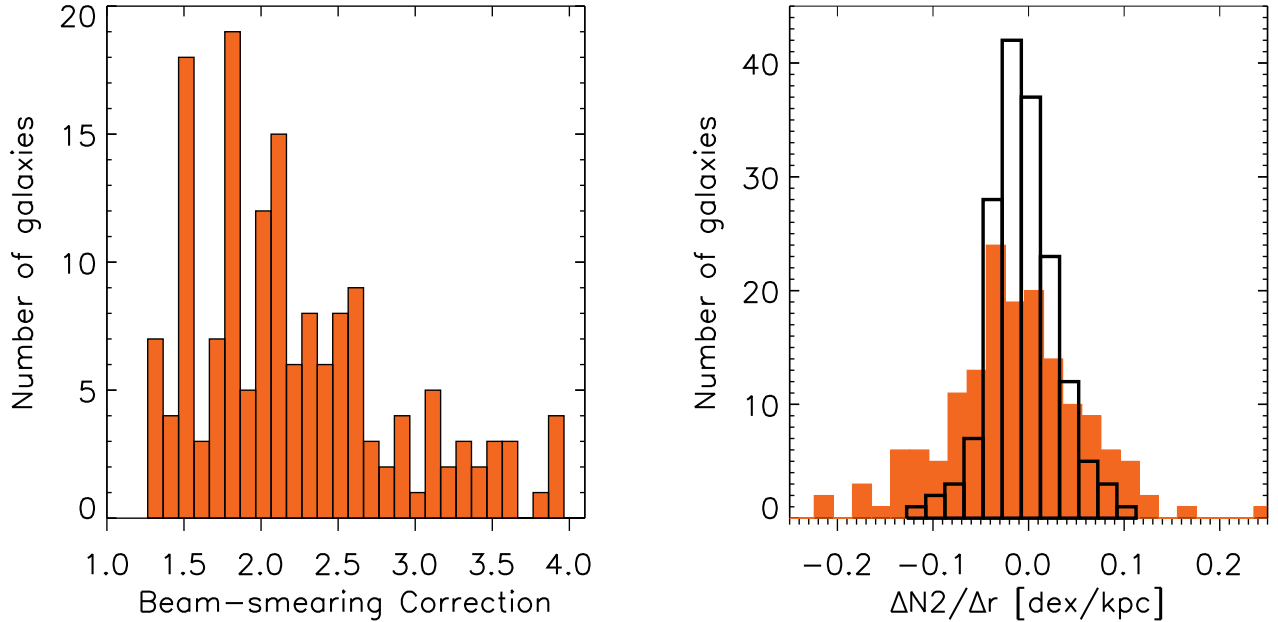


FIG. 8.— (*Left*) Histogram of the required beam-smearing corrections for the 180 gradients measured in the KMOS^{3D} sample. The corrections are based on the relevant seeing, and the effective radius and inclination of the targets as derived from GALFIT modelling of the F160W CANDELS photometry. (*Right*) Histograms of observed (black) and intrinsic (filled orange) $\Delta N_2/\Delta r$ gradients.

of high redshift star-forming galaxies, this might be too strong of a simplification. In future work we will explore this in more detail with hydrodynamical simulations.

4.2. Gradients in Interacting Systems

In the local Universe, galaxies undergoing an interaction show flattened gradients (e.g. Kewley et al. 2010; Rupke et al. 2010b; Sánchez et al. 2014). This is expected from simulations, as the interaction causes inflows of metal-poor gas from the outskirts into the galaxy centre (Rupke et al. 2010b; Perez et al. 2011). At high redshift, the situation remains unclear. For their samples of lensed galaxies, Jones et al. (2013) and Leethochawalit et al. (2015) report shallower gradients for the merging or dynamically disturbed systems. Similarly, four out of the seven galaxies with a positive abundance gradient in the seeing-limited MASSIV sample at $z \sim 1.2$ are interacting (Queyrel et al. 2012). In contrast, Cresci et al. (2010) and Troncoso et al. (2014) find positive gradients in rotating disks, which would require the inflow of metal-poor gas into the galaxy centre through minor mergers, cold flows or violent disk instabilities (e.g. Dekel & Burkert 2014; Zolotov et al. 2015).

Within the KMOS^{3D} sample, we find only a handful of galaxies with a statistically significant positive gradient. Wisnioski et al. (2015) define a sample of pairs based on a search for companions in the 3D-HST catalog with a velocity separation of 500 km/s, expected to fall within the IFU of the primary target - this roughly translates to a projected separation of $1.5''$ or 12 kpc. There are 39 such targets in the sample, for which we can recover 18 gradients. Additionally, one of the companion galaxies is sufficiently detected and extended for an additional gradient measurement. We also extend our definition of pairs to the more commonly used projected separation < 50 kpc (restricted to spectroscopic redshifts), and classify 4 additional KMOS^{3D} targets with a gradient measurement as interacting, for a total sample of 23. Figure 9 compares the 23 gradients in interacting systems to the full KMOS^{3D} sample. The median gradient of 0.005 ± 0.008 (standard deviation 0.036) is not significantly flattened compared to the full median of -0.008 ± 0.003 (standard deviation 0.035). A Kolmogorov-Smirnov test indicates a likelihood of 29% that both sets originate from the same parent sample. When first applying the beam-smearing correction, we find 0.011 ± 0.016 (standard deviation 0.08) for the pairs and -0.015 ± 0.008 (standard deviation 0.10) for the full sample, with a KS-test likelihood of 31%.

4.3. Evolution of Metallicity Gradients with Cosmic Time

Figure 10 compiles available abundance gradients from the literature as a function of redshift. For this compilation, we convert our gradients in $[\text{N II}]/\text{H}\alpha$ ($\Delta\text{N2}/\Delta r$) into abundance gradients $\Delta Z/\Delta r$ by multiplying with a factor 0.57 in accordance with the linear conversion by Pettini & Pagel (2004). In the local Universe, the median gradient for a sample of isolated spiral galaxies from Rupke et al. (2010b) and Zaritsky et al. (1994) is -0.04 ± 0.05 dex/kpc. The open square symbol shows the flatter gradient of -0.0165 dex/kpc for galaxies currently undergoing an

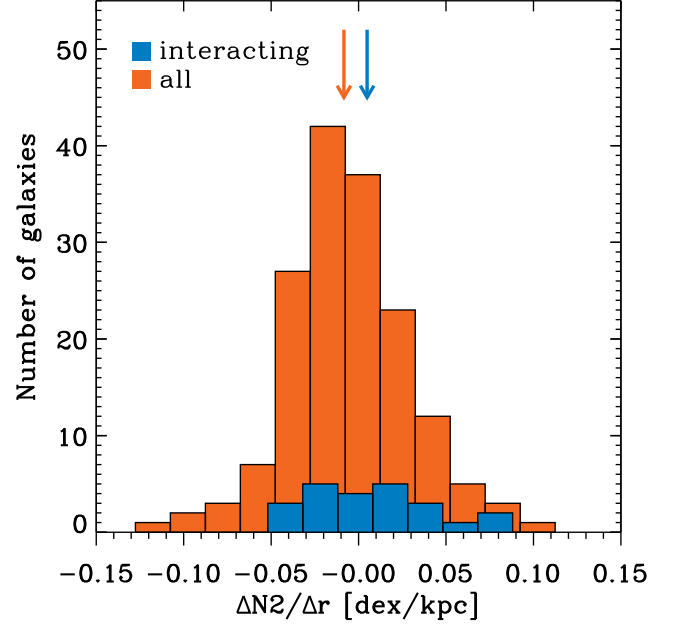


FIG. 9.— Comparison of the full sample of KMOS^{3D} $\Delta\text{N2}/\Delta r$ gradients in orange versus the 23 interacting systems in blue. The difference is not statistically significant. The orange and blue arrows indicate the median gradient of -0.008 ± 0.003 and 0.005 ± 0.008 for the full sample and pairs respectively. Given the uncertainties, the interacting systems do not show significantly flatter gradients.

interaction (Rupke et al. 2010b). These studies use the R_{23} metallicity indicator of $([\text{O II}] \lambda\lambda 3727, 3729 + [\text{O III}] \lambda 4959 + [\text{O III}] \lambda 5007)/\text{H}\beta$. Rupke et al. (2010b) compare their results to abundances calculated from $[\text{N II}] \lambda 6583/[\text{O II}] \lambda\lambda 3727, 3729$ and $([\text{O III}] \lambda 5007)/\text{H}\beta)/([\text{N II}] \lambda 6583/\text{H}\alpha)$ (the O3N2 indicator), and find no systematic difference after calibrating the different indicators following Kewley & Ellison (2008). At high redshift, abundance gradients are almost exclusively derived from the $[\text{N II}]/\text{H}\alpha$ ratio, since it only requires observations in one band and is insensitive to dust correction. Jones et al. (2013) and Leethochawalit et al. (2015) additionally measure O3N2 for a subset of their samples and find consistent results. The grism data from Jones et al. (2015), and the $z > 3$ sample from Troncoso et al. (2014) do not have access to $\text{H}\alpha$, and use the R23 diagnostic instead. For the literature sample in Figure 10, the star symbols correspond to lensed galaxies with AO, diamonds are non-lensed galaxies with AO, and squares show seeing-limited data. Open symbols show interacting or disturbed systems, as defined in the various papers from kinematic classification, visual inspection of broad-band imaging, or a combination of both. For KMOS^{3D}, we use the pair classification as described above.

With the availability of larger samples, the steep abundance gradients found in lensed galaxies by Jones et al. (2013) and Yuan et al. (2011) that spurred much of the initial debate on the feasibility of gradient measurements in field galaxies, as well as the idea of a strong flattening of gradients over cosmic time, now seem more

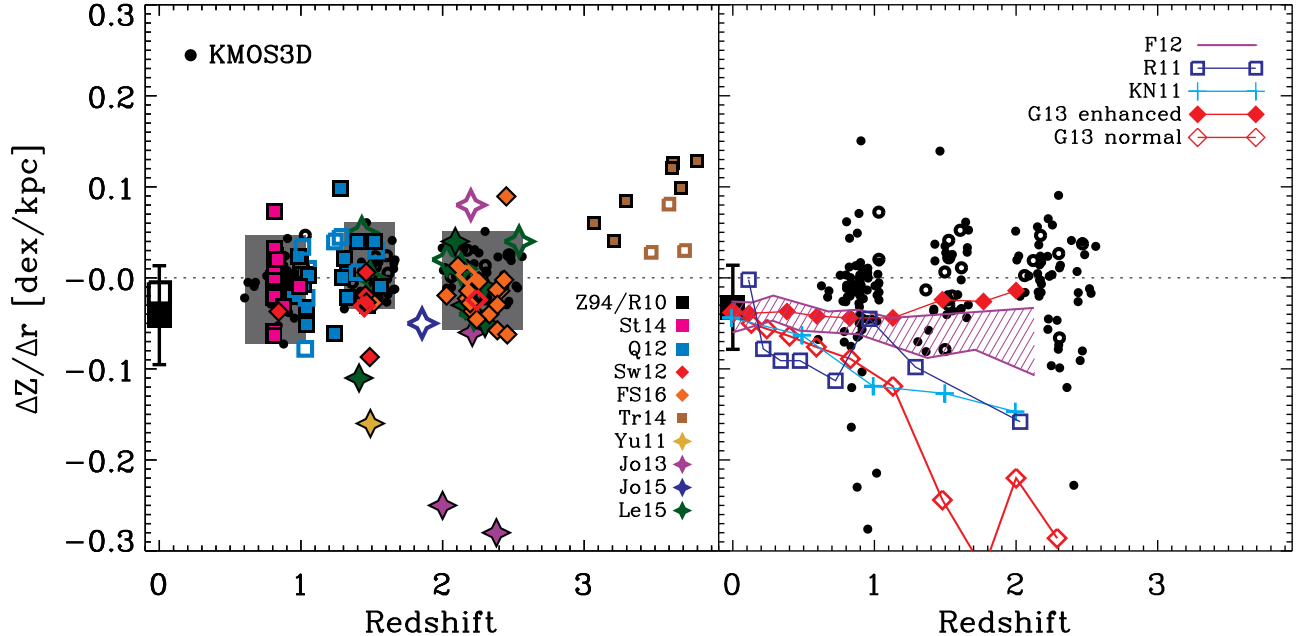


FIG. 10.— The left panel compiles the KMOS^{3D} results with all the abundance gradients measured at high redshift from the literature. Open symbols refer to interacting or disturbed systems, as defined in the various papers. In the local Universe, the filled/open black square shows the median gradient for a sample of isolated spiral galaxies (Rupke et al. 2010b; Zaritsky et al. 1994) and interacting spiral systems (Rupke et al. 2010b) respectively. The star symbols refer to gravitationally lensed galaxies (Yuan et al. 2011; Jones et al. 2013, 2015; Leethochawalit et al. 2015); diamonds have AO-assisted data (Swinbank et al. 2012, Förster Schreiber et al. 2016 (in prep.)); squares have only seeing-limited data (Queyrel et al. 2012; Stott et al. 2014; Troncoso et al. 2014). KMOS^{3D} data is represented by small black open or closed circles. To aid visibility, the grey background squares indicate the spread of the KMOS^{3D} measurements in the three redshift intervals. In the *right* panel, the KMOS^{3D} galaxies have been corrected for the effect of beam-smearing. We additionally plot different theoretical predictions for the time evolution of abundance gradients from Pilkington et al. (2012) based on simulations in Rahimi et al. (2011); Kobayashi & Nakasato (2011); Few et al. (2012); Gibson et al. (2013).

like outliers. These lensed galaxies have lower stellar masses than typically targeted in the other surveys (the three steepest gradients in Figure 10 were measured for $\log(M_*/M_\odot) = 9.7$ and 9.9 (Jones et al. 2013) and a dynamical mass of $10.42 M_\odot$ (Yuan et al. 2011)), but there is still overlap. The KMOS^{3D} survey for example probes down to $\log(M_*/M_\odot) = 9.7$, as can be seen in Figure 6. Also, the larger study of lensed galaxies in Leethochawalit et al. (2015) did not uncover any additional steep gradients.

In the right panel of Figure 10, the KMOS^{3D} gradients are corrected for beam-smearing. We can not extend this correction to the literature data, since we do not always know the FWHM, effective radius and inclination of the targets. Overplotted are models for the evolution of gradients over time from Rahimi et al. (2011); Kobayashi & Nakasato (2011); Few et al. (2012) and Gibson et al. (2013) as derived by Pilkington et al. (2012). The red open and filled diamonds compare the normal and enhanced feedback mode for the same simulation from Gibson et al. (2013). There is large scatter in the observed gradients, and remaining uncertainty with regards to the beam-smearing correction, but based on these models, the scarcity of gradients much steeper than -0.1 dex/kpc at $z \sim 2$ suggests the need for strong feedback.

This work takes advantage of the large sample size and wide redshift coverage of the KMOS^{3D} near-IR IFU survey to infer the integrated metal abundance and abundance gradient of star-forming galaxies at the peak of cosmic star formation activity between $z = 0.6 - 2.6$ from the $[\text{N II}]/\text{H}\alpha$ flux ratio. Such measurements can shed crucial light on the main physical processes driving the assembly of stellar mass and heavy elements in the Universe. Our main conclusions are as follows:

- Our mass-metallicity relation confirms previous literature results. Excluding AGN reduces sensitivity at the high-mass end, but does not significantly change the relation at lower masses ($\log(M_*/M_\odot) \lesssim 10.8$). We do not find a significant effect of SFR, sSFR or galaxy environment on the observed $[\text{N II}]/\text{H}\alpha$ ratios.
- A thorough analysis of the effect of beam-smearing in seeing-limited data at high redshift shows that one can recover between 30% and 70% of the intrinsic abundance gradient, depending mostly on the ratio of the FWHM of the data to the effective radius, and to a lesser extent on galaxy inclination.
- The majority of observed $[\text{N II}]/\text{H}\alpha$ gradients are flat. Cosmological simulations suggest the need for strong feedback to create flat abundance gradients at high redshift. No statistically significant correlations with stellar, kinematic or structural galaxy

properties are found.

A number of factors could cause intrinsically flat abundance gradients. One effect is the radial inflow of metal-poor gas and metal-mixing caused by a merger. In the local Universe, this has been established to cause shallower gradients, both observationally and in simulations (Rupke et al. 2010a,b). At high redshift, the merger fraction remains highly debated with most observational results pointing to consistent or higher fractions than found locally (e.g. Lotz et al. 2008; Man et al. 2012; López-Sanjuan et al. 2014; Lackner et al. 2014). However, not all high- z surveys systematically find shallower gradients for disturbed systems, as discussed in Sections 4.2 and 4.3.

A second important physical process is feedback, which can blow out metal-rich gas from the centre and redistribute it at larger radii. The ubiquitous presence of both AGN and star formation driven feedback is well established at high redshift, especially at high mass and star formation surface density (e.g. Newman et al. 2012; Genzel et al. 2014).

Aside from these physical effects, one has to take into account the various reasons why the *observed* gradient could be flat. Beam-smearing is the most important factor here, and while we have attempted to address this, significant uncertainties remain. Secondly, in analogy with local SFGs, the fraction of warm diffuse ionised gas

increases in radial apertures further out in the galaxy disks, and could strongly bias the measured line ratios there (Yuan et al. 2011). Adaptive-optics assisted data reaches the resolution of massive star-forming clumps at $z \sim 2$ and could thus minimise this effect, but the uncorrected seeing-limited halo still contributes to the beam-smearing correction (Figure 7). Space-based grism data is more suited, but struggles with sensitivity.

Similar biases are introduced when shock fronts alter the photo-ionization conditions of the interstellar medium (e.g. Kewley et al. 2013; Newman et al. 2014). These biases affect any abundance indicator based on rest-frame optical emission line ratios, but have especially significant effects for the N2 indicator. Spatially resolved coverage of multiple lines would allow better rejection of spatial pixels influenced by AGN or shock ionisation based on the BPT diagram.

This analysis has highlighted the need for high-quality data with sufficient spatial resolution, S/N and wavelength coverage to disentangle the physical processes driving the observed line ratios and to pin down the intrinsic abundance gradients. Our understanding of the global and resolved metallicity distributions among and within SFGs, as well as their cosmic evolution, will benefit greatly from the added resolution and sensitivity of JWST and the ELTs.

APPENDIX

MF and DW acknowledge the support of the Deutsche Forschungs Gemeinschaft (DFG) via Project WI 3871/1-1.

REFERENCES

- Andrews, B. H., & Martini, P. 2013, *ApJ*, 765, 140
 Anglés-Alcázar, D., Davé, R., Özel, F., et al. 2014, *ApJ*, 782, 84
 Aumer, M., White, S. D. M., & Naab, T. 2014, *MNRAS* 441, 3679
 Baldwin, J. A., Phillips, M. M., Terlevich, R., 1981, *PASP*, 93, 5
 Belli, S., Jones, T., Ellis, R., et al. 2013, *ApJ*, 772, 141
 Blanc, G. A.; Kewley, L. J., Vogt, F. P. A, et al. 2015, *ApJ*, 798, 99
 Bouché, N., Dekel, A., Genzel, R., et al. 2010, *ApJ*, 718, 1001
 Bournaud, F., Perret, V.; Renaud, F., et al. 2014, *ApJ*, 780, 57
 Brammer, G. B., van Dokkum, P. G., Franx, M., et al. 2012, *ApJS*, 200, 13
 Bresolin, F., Kennicutt, R. C., & Ryan-Weber, E. 2012, *ApJ*, 750, 122
 Brooks, A. M., Governato, F., Booth, C. M., et al. 2007, *ApJ*, 655, L17
 Bruzual, G. & Charlot, S. 2003, *MNRAS*, 344, 1000
 Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, *ApJ*, 533, 682
 Ceverino, D., Sanchez Almeida, J., Muñoz-Tuñón, C., et al. 2015, *arXiv:1509.02051*
 Chabrier, G. 2003, *PASP*, 115, 763
 Chiappini, C., Matteucci, F., & Romano, D. 2001, *ApJ*, 554, 1044
 Cid Fernandes, R., Stasińska, G., Mateus, A., & Vale Asari, N. 2011, *MNRAS*, 413, 1687
 Cresci, G., Hicks, E. K. S., Genzel, R., et al. 2009, *ApJ*, 697, 115
 Cresci, G., Mannucci, F., Maiolino, R., et al. 2010, *Nature*, 467, 811
 Coil, A. L., Aird, J., Reddy, N., et al. 2015, *ApJ*, 801, 35
 Cullen, F., Cirasuolo, M., McLure, R. J., et al. 2014, *MNRAS*, 440, 2300
 Dalcanton, J. J., Yoachim, P., & Bernstein, R. A. 2004, *ApJ*, 608, 189
 Davé, R., Finlator, K., & Oppenheimer, B. D. 2011, *MNRAS*, 416, 1354
 Davé, R., Finlator, K., & Oppenheimer, B. D. 2012, *MNRAS*, 421, 98
 Davies, R., Förster Schreiber, N. M., Cresci, G., et al. 2011, *ApJ*, 741, 69
 Davies, R., Agudo Berbel, A., Wierorrek, E., et al. 2013, *A&A*, 558, A56
 Dekel, A., Sari, R., Ceverino, D. et al. 2009, *ApJ*, 703, 785
 Dekel, A., & Burkert, A. 2014, *MNRAS*, 438, 1870
 Ellison, S. L., Patton, D. R., Simard, L., & McConnell, A. W. 2008, *ApJ*, 672, L107
 Ellison, S. L., Simard, L., Cowan, N. B., et al. 2009, *MNRAS*, 396, 1257
 Erb, D. K., Shapley, A. E., Pettini, M., Steidel, C. C., Reddy, N. A., Adelberger, K. L. 2006, *ApJ*, 644, 813
 Few, C. G., Gibson, B. K., Courty, S., et al. 2012, *A&A*, 547, 63
 Finlator, K., & Davé, R. 2008, *MNRAS*, 385, 2181
 Förster Schreiber, N. M., Genzel, R., Bouché, N., et al. 2009, *ApJ*, 706, 1364
 Förster Schreiber, N. M., Genzel, R., Newman, S. F., et al. 2014, *ApJ*, 787, 38
 Fossati, M., Wilman, D. J., Fontanot, F., et al. 2015, *MNRAS*, 446, 2582
 Fu, J., Hou, J. L., Yin, J., et al. 2009, *ApJ*, 696, 668
 Genel, S., Genzel, R., Bouché, N., et al. 2008, *ApJ*, 688, 789
 Genel, S., Naab, T., Genzel, R., et al. 2012, *ApJ*, 745, 11
 Genzel, R., Förster Schreiber, N. M., Rosario, D., et al. 2014, *ApJ*, 796, 7
 Gibson, B. K., Pilkington, K., Brook, C. B. et al. 2013, *A&A*, 554, 47
 Goddard, Q. E., Bresolin, F.; Kennicutt, R. C et al. 2011, *MNRAS*, 412, 1246
 Grasshorn Gebhardt, H. S., Zeimann, G. R., Ciardullo, R. et al. 2015, *arXiv:1511.08243*

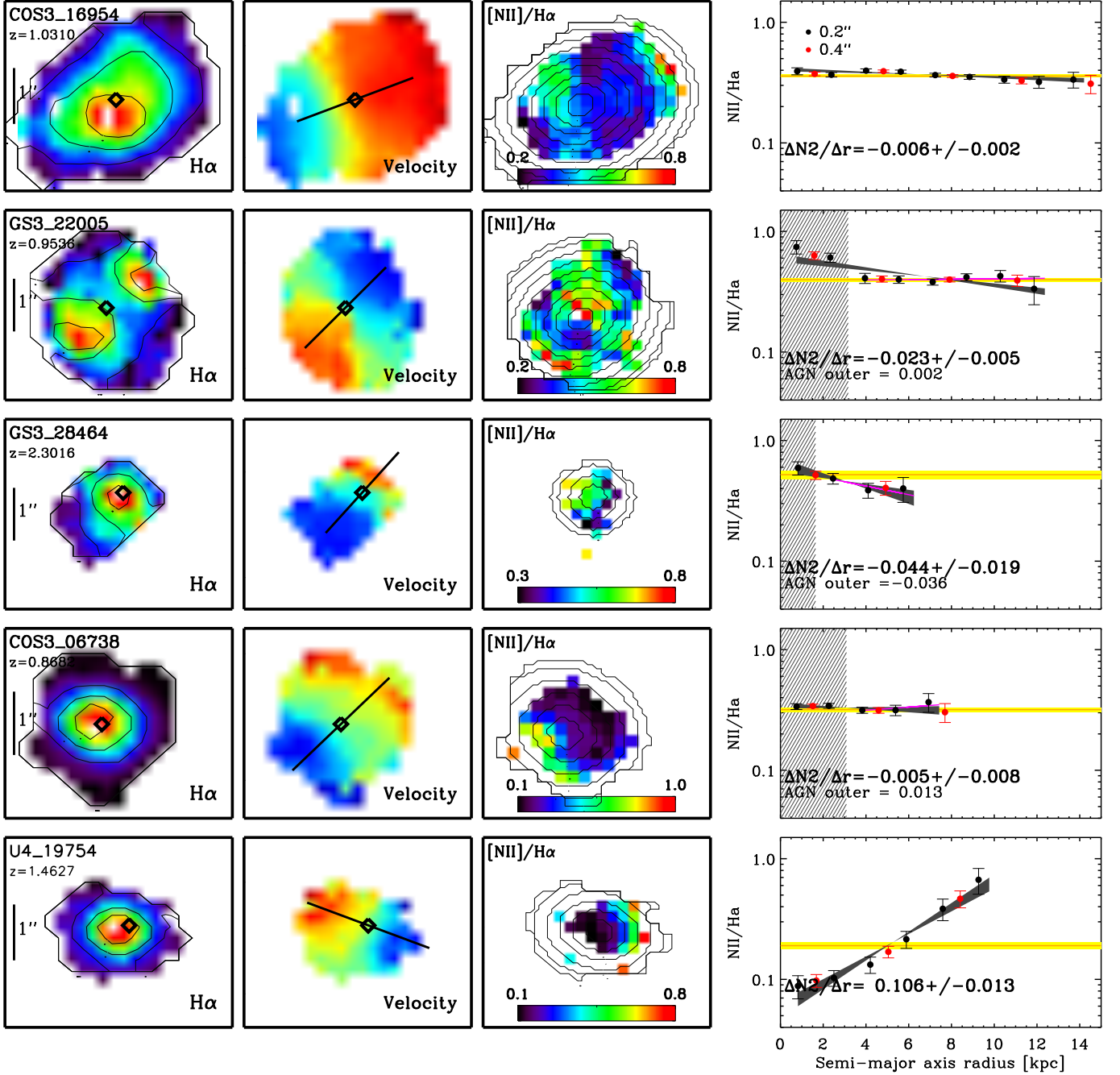


FIG. 11.— Illustration of how abundance gradients are measured for five example galaxies. The three panels on the left show the $H\alpha$, velocity and $[N II]/H\alpha$ maps respectively, where in the latter only pixels with $S/N > 3$ in $[N II]$ are shown. In each panel, the black open diamond corresponds to the continuum centre. The $H\alpha$ map additionally shows the $H\alpha$ contours, and the kinematic position angle is shown on the velocity map. Based on the centre, PA and ellipticity of the outer $H\alpha$ contours, we create one-pixel wide elliptical apertures as shown on top of the $[N II]/H\alpha$ map in the third column. The chosen ellipticity does not significantly affect the measured gradient. The rightmost panels show the $[N II]/H\alpha$ ratios as a function of semi-major axis radius for these apertures. The red data points correspond to the wider 2 pixel or 0.4'' apertures. The integrated $[N II]/H\alpha$ ratio and its uncertainty are indicated by the orange horizontal line and yellow band. The dark grey shaded band is the best-fit gradient and uncertainty. The middle 3 galaxies are each identified as an AGN. In this case we exclude the inner one or two apertures as indicated by the vertical grey shaded region, and refit the remaining outer data points (magenta line). For the third and fourth galaxy, both fits are consistent. The second row shows an example where the $[N II]/H\alpha$ ratios in the centre are significantly elevated due to the AGN, such that the outer gradient is much flatter.

Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, *ApJS*, 197, 35
 Henry, A., Scarlata, C., Domínguez, A., et al. 2013, *ApJ*, 776, L27
 Hughes, T. M., Cortese, L., Boselli, A., et al. 2013, *A&A*, 550, 115
 Jones, T., Ellis, R. S., Jullo, E., et al. 2010, *AJ*, 725, L176
 Jones, T., Ellis, R. S., Richard, J., et al. 2013, *ApJ*, 765, 48
 Jones, T., Wang, X., Schmidt, K. B. et al. 2015, *AJ*, 149, 107

Kacprzak, G. G., Yuan, T.-T., Nanayakkara, T., et al. 2015, *ApJ*, 802, L26
 Kewley, L. J., Dopita, M. A., Sutherland, R. S., et al. 2001, *ApJ*, 556, 121
 Kewley, L. J., Ellison, S. L. 2008, *ApJ*, 681, 1183
 Kewley, L. J., Rupke, D., Zahid, J. H. et al. 2010, *ApJ*, 721, 48

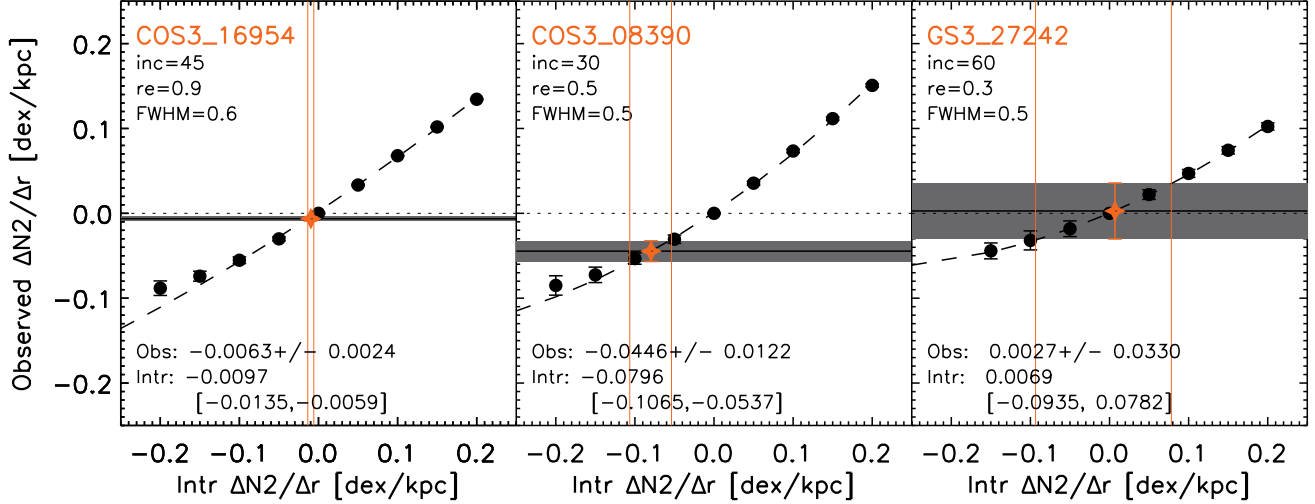


FIG. 12.— This figure provides three examples of the beam-smearing correction and intrinsic metallicity gradient. In each panel, the horizontal grey band and black line show the observed gradient and its uncertainty. The black circles correspond to the gradient we would have observed for a range of intrinsic gradients from -0.2 to 0.2 dex/kpc, given the known seeing, inclination and effective radius of the target. We fit a polynomial to this correlation, and invert it to derive the intrinsic gradient and uncertainties which corresponds to the gradient we actually observed in the data. This is shown by the orange star and orange vertical lines. The observed and intrinsic gradients are noted in [dex/kpc] in the bottom of each panel. In the right panel, the relation between observed and intrinsic gradient is very flat due to the small effective radius and high inclination of the target, requiring a large correction. This additionally causes large uncertainties on the intrinsic gradient compared to the observed uncertainties.

Kewley, L. J., Dopita, M. A., Leitherer, C. et al. 2013, *ApJ*, 774, 100
Kobayashi, C., & Nakasato, N. 2011, *ApJ*, 729, 16
Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, *ApJS*, 197, 36
Kriek, M., van Dokkum, P. G., Labbé, I., et al. 2009, *ApJ*, 700, 221
Kriek, M., Shapley, A. E., Reddy, N. A., et al. 2015, *ApJS*, 218, 15
Kulas, K. R., McLean, I. S., Shapley, A. E., et al. 2013, *ApJ*, 774, 130
Lackner, C. N., Silverman, J. D., Salvato, M., et al. 2014, *AJ*, 148, 137
Leethochawalit, N., Jones, T. A., Ellis, R. S. et al. 2015, *arXiv:1509.01279*
Lilly, S. J., Carollo, C. M., Pipino, A., Renzini, A., & Peng, Y. 2013, *ApJ*, 772, 119
López-Sanjuan, C., Cenarro, A. J., Hernández-Monteagudo, C., et al. 2014, *A&A*, 564, A127
Lotz, J. M., Davis, M., Faber, S. M., et al. 2008, *ApJ*, 672, 177
Lu, Y., Blanc, G. A., Benson, A. 2015, *ApJ*, 808, 129L
Lutz, D., Poglitsch, A., Altieri, B., et al. 2011, *A&A*, 532, A90
Magdis, G. E., Bureau, M., Stott, J. P., et al. 2016, *MNRAS*, 456, 4533
Man, A. W. S., Toft, S., Zirm, A. W., et al. 2012, *ApJ*, 744, 85
Magnelli, B., Popesso, P., Berta, S., et al. 2013, *A&A*, 553, A132
Mannucci, F., Cresci, G., Maiolino, R., Marconi, A., & Gnerucci, A. 2010, *MNRAS*, 408, 2115
Mollá, M., Díaz, A. I. 2005, *MNRAS*, 358, 521
Momcheva, I. G., Brammer, G. B., van Dokkum, P. G., et al. 2015, *arXiv:1510.02106*
Nelson, E. J., van Dokkum, P. G., Förster Schreiber, N. M., et al. 2015, *arXiv:1507.03999*
Newman, S. F., Genzel, R., Förster-Schreiber, N. M. et al. 2012, *ApJ*, 761, 43
Newman, S. F., Buschkamp, P., Genzel, R. et al. 2014, *ApJ*, 781, 21
Oppenheimer, B. D., & Davé, R. 2008, *MNRAS*, 387, 577
Peeples, M. S., & Shankar, F. 2011, *MNRAS*, 417, 2962
Perez, J., Michel-Dansac, L., & Tissera, P. B. 2011, *MNRAS*, 417, 580
Pérez-Montero, E. 2014, *MNRAS*, 441, 2663
Pettini, M. & Pagel, B. E. J. 2004, *MNRAS*, 348, L59

Pilkington, K., Few, C. G., Gibson, B. K., et al. 2012, *A&A*, 540, 56
Queyrel, J., Contini, T., Kissler-Patig, M., et al. 2012, *A&A*, 539, A93
Rahimi, A., Kawata, D., Allende Prieto, C. et al. 2011, *MNRAS*, 415, 1469
Rudie, G. C., Steidel, C. C., Trainor, R. F. et al. 2012, *ApJ*, 750, 67
Rupke, D. S. N., Kewley, L. J., & Barnes, J. E. 2010, *ApJ*, 710, 156L
Rupke, D. S. N., Kewley, L. J., & Chien, L.-H. 2010, *ApJ*, 723, 1255
Sánchez, S. F., Rosales-Ortega, F. F., Iglesias-Prado, J. et al. 2014, *A&A*, 563, 49
Sánchez-Blázquez, P., Rosales-Ortega, F. F., Mndez-Abreu, J. et al. 2014, *A&A*, 570, 6
Sánchez Almeida, J., Morales-Luis, A. B., Muñoz-Tuñón, C., et al. 2014, *ApJ*, 783, 45
Sanders, R. L., Shapley, A. E., Kriek, M., et al. 2015, *ApJ*, 799, 138
Scudder, J. M., Ellison, S. L., & Mendel, J. T. 2012, *MNRAS*, 423, 2690
Silverman, J. D., Kashino, D., Sanders, D. et al. 2015, *ApJS*, 220, 12
Sharples, R., Bender, R., Agudo Berbel, A., et al. 2013, *Msngr*, 151, 21
Shapley, A. E., Reddy, N. A., Kriek, M., et al. 2015, *ApJ*, 801, 88
Shimakawa, R., Kodama, T., Tadaki, K., et al. 2015, *MNRAS*, 448, 666
Skelton, R. E. and Whitaker, K. E. and Momcheva, I. G., et al. 2014, *ApJS*, 214, 24
Spitoni, E., Calura, F., Matteucci, F., et al. 2010, *A&A*, 514, A73
Stasińska, G., Cid Fernandes, R., Mateus, A., et al. 2006, *MNRAS*, 371, 972
Steidel, C. C., Shapley, A. E., Pettini, M., Adelberger, K. L., Erb, D. K., Reddy, N. A., & Hunt, M. P. 2004, *ApJ*, 604, 534
Steidel, C. C., Rudie, G. C., Strom, A. L., et al. 2014, *ApJ*, 795, 165
Storey, P. J., & Zeppen, C. J. 2000, *MNRAS*, 312, 813
Stott, J. P., Sobral, D., Bower, R., et al. 2013, *MNRAS*, 436, 1130
Stott, J. P., Sobral, D., Swinbank, A. M., et al. 2014, *MNRAS*, 443, 2695
Swinbank, A. M., Sobral, D., Smail, I., et al. 2012, *MNRAS*, 426, 935

TABLE 3
TARGET PROPERTIES.

#	z	$\log(M_*/M_\odot)$	$\text{SFR}_{\text{H}\alpha}^{\text{a}}$	$[\text{N II}]/\text{H}\alpha^{\text{b}}$	$\Delta N2/\Delta r$	beam-smearing
			$[\text{M}_\odot/\text{yr}]$		$[\text{dex/kpc}]$	
1	0.758	10.66	5.5	0.41 ± 0.03	0.004 ± 0.016	1.3
2	0.775	10.94	0.9	< 0.11		
3	0.777	10.35	24.7	0.44 ± 0.03	-0.013 ± 0.015	3.4
4	0.787	10.78	4.6	0.56 ± 0.03	-0.038 ± 0.013	3.3
5	0.787	10.51	21.0	0.41 ± 0.03	-0.050 ± 0.014	2.1

^{NOTE} This Table is published in its entirety in the electronic edition of the Astrophysical Journal. A portion is shown here for guidance regarding its form and content.

^aH α -based SFR corrected for dust extinction using the SED-derived reddening and accounting for additional extinction of the nebular lines following Wuyts et al. (2013)

^bIntegrated $[\text{N II}]/\text{H}\alpha$ ratio

^cObserved gradient in $[\text{N II}]/\text{H}\alpha$, excluding the nuclear region for AGN-flagged targets. Reported only when measurable.

^dMultiplicative correction factor for the N2 gradient.

Tacconi, L. J., Neri, R., Genzel, R., et al. 2013, ApJ, 768, 74
Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. 2004, ApJ, 613, 898
Treu, T., Schmidt, K. B., Brammer G. B., et al. 2015, arXiv1509.00475
Troncoso, P., Maiolino, R., Sommariva, V., et al. 2014, A&A, 563, 58
Valentino, F., Daddi, E., Strazzullo, V. et al. 2015, ApJ, 801, 132
van der Wel, A., Bell, E. F., Häussler, B., et al. 2012, ApJS, 203, 24
van Dokkum, P. G., Leja, J., Nelson, E. J., et al. 2013, ApJ, 771, L35
van Zee, L., Salzer, J. J., Haynes, M. P., et al. 1998, AJ, 116, 2805
Wiersma, R. P. C., Schaye, J., & Smith, B. D 2009, MNRAS, 393, 99
Wisnioski, E., Förster Schreiber, N. M., Wuyts, S., et al. 2015, ApJ, 799, 209
Wuyts, E., Kurk, J., Förster Schreiber, N. M., et al. 2014, ApJ, 789, L40

Wuyts, S., Förster Schreiber, N. M., Lutz, D., et al. 2011, ApJ, 738, 106
Wuyts, S., Förster Schreiber, N. M., Genzel, R., et al. 2012, ApJ, 753, 114
Wuyts, S., Förster Schreiber, N. M., Nelson, E. J., et al. 2013, ApJ, 779, 135
Yuan, T.-T., Kewley, L. J., Swinbank, A. M., et al. 2011, ApJ, 732, L14
Yuan, T.-T., Kewley, L. J., & Rich, J. 2013, ApJ, 767, 106
Zahid, H. J., Kewley, L. J., & Bresolin, F. 2011, ApJ, 730, 137
Zahid, H. J., Kashino, D., Silverman, J. D., et al. 2014, ApJ, 792, 75
Zaritsky, D., Kennicutt, R. C., Jr., & Huchra, J. P. 1994, ApJ, 420, 87
Zolotov, A., Dekel, A., Mandelker, N., et al. 2015, MNRAS, 450, 2327